Methodology in Medical Research

The Need for Controlled Clinical Studies

ROBERT A. HALL, M.D., Agnew

• Too many medical researchers vitiate their work by ignoring the problem of uncontrolled variables. They therefore publish clinical impressions "dressed up" in scientifically meaningless numbers. A prototypical example of this practice is contrasted with a controlled study, each employing the same (small) number of patients. It is shown how the use of controls can convert a meaningless experiment into one that has assessable scientific significance.

A survey of current literature revealed that in only 21 of 100 articles studied were adequately controlled experimental conditions employed.

Since they usually deal with very complex systems, it is urged that medical researchers exercise more scientific rigor with regard to control problems.

A SINE QUA NON of the scientific method is the principle of control, namely, that in order to evaluate the effect of one variable factor in an experimental situation, one must hold constant all other variable factors.

Among physical scientists, the application of the principle of control is almost automatic, ingrained, inviolable. In published medical research, however, a regrettably frequent lack of attention to this principle may be noted.

Medical researchers, notably those who publish reports of work in clinical fields, too often vitiate the fruits of their labors by not maintaining adequate controls. They lose the opportunity of bringing forth solid scientific facts; instead they too often produce something little removed from "clinical impression." This is not to detract from the importance of the clinical impression, which is most necessary and valuable, but as a stimulus for systematic scientific research by which it can be substantiated or disproven, not as an end in itself. Too frequently such impressions, window-dressed with a few numbers, serve to stimulate great and often tragic optimism on the part of the unwary. It is a medical aphorism that "every new drug works—for a while."

A typical, familiar situation is one such as this: Dr. X, a clinician, studies the efficacy of a new drug or procedure on a group of patients who have some more or less well-defined condition. He sets up criteria for evaluating their response. He then reports something like this:

	No. Patients	Per Cent
Improved	16	80
Unimproved	4	20
Total	20	100

What do these results mean to us? We see that 80 per cent of his patients improved. It looks like a promising treatment, *maybe*. But why maybe?

Well, we ask, how many of these patients would have improved without any treatment? How much responsibility for the result can we attribute to variations in the auxiliary treatment, to the treatment factors other than the one on which Dr. X has focused his attention? How can we estimate the possible psychotherapeutic effect on these patients of merely having been given a "new treatment," as well as the impact of Dr. X's personality? Can Dr. X feel sure that his evaluation of improvement was unaffected by his enthusiasm for the treatment? One could go on listing many more uncontrolled variables, any one or combination of which may have been the actual cause of the observed results.

The treatment under study may have been beneficial, harmful, or ineffective. Unfortunately, we will never know in a factual scientific sense—at least not from the data given us.

SIMPLE CLINICAL CONTROLS

How could Dr. X have put us on more solid ground? He could have used one or both of two commonly employed control methods:

1. *Control Group Method.* Make a random distribution of homogeneous patients into two groups, approximately equal in size. As far as possible, treat and evaluate each group *exactly alike,* except in respect to one thing: the treatment in question. Placebos are often essential to provide the latter. After obtaining the results, statistically analyze the probability of the observed differences between groups being due to chance alone.

From Agnews State Hospital, California State Department of Mental Hygiene, Agnew, and the Psychiatric Section, Santa Clara County Hospital, San Jose.

Submitted January 28, 1955.

2. *Longitudinal Method.* Use each patient as his own control. Observe the patient during a control period, then during alternate periods of receiving drug and placebo. Evaluate results statistically when possible. Obviously, this method can be best applied in conditions which are more or less chronic and stable.

It is often possible to set up either method to include the double blind control technique. This involves designing the experiment so that neither patients nor observers know which patients are "controls" and which are "experimentals." Rarely can one feel that the psychological factors in subject and observer are controlled if this is not done.

In regard to the control group method, when it is difficult to assemble or handle large numbers of cases, the great temptation is to use few or none as controls. But even with few total patients, it is worth while to follow the method. Let us see what Dr. X might have done with the same amount of clinical material and very little extra effort.

Suppose Dr. X had made a random distribution of his 20 cases into two equal groups. In every respect he could think of, he had handled each group exactly alike, except that the patients of the control group received a placebo instead of the treatment under study. He had set up a system whereby he did not know himself to which group a particular patient belonged until after he had finished evaluating the whole series. He then reported as follows:

	Treated		Controls		Total	
	No.	Per Cent	No.	Per Cent	No.	Per Cent
Improved	5	50	0	0	5	25
Unimproved	5	50	10	100	15	75
Total	10	100	10	100	20	100

This time we see that only one-half of the treated group was helped. At first glance we are not as much impressed as we were with the data as previously reported. But then we look at the control group and find that *none* of them improved.

Is it possible that Dr. X happened to give the drug to a group of patients who were going to improve regardless of the treatment? Could the patients in the control group have been the sicker ones? In other words, could the observed differences between the groups have been due to chance alone?

We use a simple statistical procedure* and discover that there is less than five chances in a hundred that such a difference could have occurred by chance alone. Now we are much more impressed by the second study. This is a definite and positive result. We know where we stand. We know to what degree the treatment deserves further trial.

These matters are very elementary. It would seem that hardly anyone would argue against the desirability of clinical investigation following controlled methods. One would hesitate to submit such a communication as this to a scientific journal were it not that disregard of the control principle is so commonplace. One can hardly pick up a clinical journal without finding one or more examples. To evaluate this latter "clinical impression," the following literature survey was pursued.

*Chi Square.

TABLE 1.—*Distribution, as to type, of 36 journals included in study*

GENERAL	**OPHTHALMOLOGY**
1. J.A.M.A.	1. A.M.A. Arch. Ophth.
2. Lancet	2. Am. J. Ophth.
3. New England J. Med.	
4. Am. Pract. & Digest Treat.	**INTERNAL MEDICINE**
5. New York J. Med.	1. Ann. Int. Med.
6. Geriatrics	2. Am. Rev. Tuberc.
7. J. Clin. Invest.	3. Am. J. Digest Dis.
8. California Med.	4. J. Allergy
	5. A.M.A. Arch. Int. Med.
SURGERY	6. Am. J. Trop. Med.
1. Am. J. Surg.	7. Circulation
2. J. Internat. Coll. Surgeons	8. J. Endocrinol.
3. West. J. Surg.	9. Am. J. Syph.
4. Surg., Gynec. & Obst.	10. Am. Heart J.
5. J. Urol.	11. A.M.A. Arch. Dermat. & Syph.
6. A.M.A. Arch. Surg.	
7. J. Neurosurg.	**PEDIATRICS**
8. J. Bone & Jt. Surg.	1. Pediatrics
	2. J. Pediat.
PSYCHIATRY AND NEUROLOGY	**ANESTHESIOLOGY**
1. J. Nerv. & Ment. Dis.	1. Anesthesiology
2. Psychosom. Med.	
3. A.M.A. Arch. Neurol. & Psychiat.	**RADIOLOGY**
	1. Am. J. Roentgenol.

A SURVEY OF CLINICAL JOURNALS

Copies of 1954 issues of the clinical journals present in the library of a teaching general hospital† were surveyed. Considered for the present study were those original articles in which controlled conditions seemed clearly indicated to substantiate the *stated conclusions* of the authors. Review papers, single case reports, and preliminary reports (which were clearly such) were eliminated.

There was no selection of the journals reviewed. The issues examined were chosen at random. The reviewing was continued until 100 pertinent articles were collected. This required covering two issues of most journals and three issues of some.‡

Thirty-six journals became involved in the study. The distribution by type of journal is shown in Table 1.

Each pertinent article was categorized as having

†Santa Clara County Hospital.

‡Some journals, and some issues of other journals present in the particular library did not become involved, however, since the issues on the shelves at the time contained no articles which fulfilled the above stated criteria.

adequate, inadequate, or no controls. The categories were defined as follows:

Adequate Controls:

Every possible attempt is made to control all variables but the one in question.

Use of a sufficiently large control group (usually about equal in number to the experimental group).

Random distribution of reasonably homogeneous cases among experimental and control groups.

"Double-blind" placebo method where possible.

Adequate exploitation, when used, of the "longitudinal control method." This includes considerable periods of evaluation of patients to observe conditions before treatment, and during drug and placebo periods. (Ideally, this approach should be combined with control group method, when possible, but this was not considered a required criteria.)

Inadequate Controls:

Failure to observe one or more indicated elements of the above criteria for adequate controls.

No Controls:

No observance whatsoever of control principles.

The results of the survey of the 100 articles were: Adequate controls, 21; inadequate controls, 30; no controls, 49. Only about one study in five was well controlled and nearly one-half were entirely uncontrolled.

A few illustrative examples will be given of each category of paper found in the survey.

Adequate Controls: A study of the value of allergic treatment for asthma and hay fever used large and nearly equal control and experimental groups. Drug and placebo administration and the clinical evaluation were done by the double blind method.

Wolf and Pinsky,[1] not only conducted a good double blind controlled experiment in a study of the therapeutic value of mephensin, but made the very provocative finding that not only were the therapeutic results indistinguishable, but drug and placebo produced identical toxic results. Such findings should set us on our guard.

Inadequate Controls: A paper concluded that people who have rage reactions show more abnormal electroencephalograms. It does not give the number of cases involved and, although controls were used, there is no indication that the electroencephalogram records were evaluated blindly.

A study of the effectiveness of intra-articular hydrocortisone in arthritis involved giving placebos to only 13 of the 88 patients involved. There was no indication that the groupings were made at random or that the results were evaluated blindly.

The response of patients with hypertension to a drug were studied, using a pretreatment control period of observation. However, no periods of placebo administration or a blind evaluation were employed.

Patients used in a study which concluded that prefrontal lobotomy was effective were selectively assigned to experimental and control groups. No placebo surgical treatment was given to the control patients, and the results were not evaluated blindly.

No Controls: One author found cortisone "effective" in rheumatic carditis after merely administering the drug to a small, selected group of patients. There was no control group or control observations, nor could the reader detect any attempt to safeguard the objectivity of the evaluation.

Isoniazid was concluded to be effective in the treatment of 65 heterogeneous psychotic patients who were not compared to any control group or to their own status while on control observation or period of placebo administration.

DISCUSSION

Biological systems are ordinarily of great complexity due to the constant operation of great numbers of independent and interrelated variables, with various complex levels of integration. Biological organizations are superimposed on chemical and physical organizations. In many respects the worker in biological fields must deal with *more* complex systems than those which confront, for example, the physicist or chemist.

A competent worker in the "pure sciences" would usually not in his wildest moment publish (or have accepted for publication) a study in which control problems were ignored. How can we hope to make sense out of our medical problems if we use less and not more scientific rigor than physical scientists?

Agnews State Hospital.

REFERENCE

1. Wolf, S., and Pinsky, R.: Effects of placebo administration and occurrence of toxic reactions, J.A.M.A., 155:339, May 22, 1954.